

Article

# Assessing Genetic Algorithm-Based Docking Protocols for Prediction of Heparin Oligosaccharide Binding Geometries onto Proteins

Samuel G. Holmes<sup>1,2</sup>  and Umesh R. Desai<sup>1,2,\*</sup> 

<sup>1</sup> Department of Medicinal Chemistry, School of Pharmacy, Virginia Commonwealth University, Richmond, VA 23298, USA; holmess@vcu.edu

<sup>2</sup> Institute for Structural Biology, Drug Discovery and Development, Virginia Commonwealth University, 800 E. Leigh Street, Suite 212, Richmond, VA 23219, USA

\* Correspondence: urdesai@vcu.edu

**Abstract:** Although molecular docking has evolved dramatically over the years, its application to glycosaminoglycans (GAGs) has remained challenging because of their intrinsic flexibility, highly anionic character and rather ill-defined site of binding on proteins. GAGs have been treated as either fully “rigid” or fully “flexible” in molecular docking. We reasoned that an intermediate semi-rigid docking (SRD) protocol may be better for the recapitulation of native heparin/heparan sulfate (Hp/HS) topologies. Herein, we study 18 Hp/HS–protein co-complexes containing chains from disaccharide to decasaccharide using genetic algorithm-based docking with rigid, semi-rigid, and flexible docking protocols. Our work reveals that rigid and semi-rigid protocols recapitulate native poses for longer chains (5→10 mers) significantly better than the flexible protocol, while 2→4-mer poses are better predicted using the semi-rigid approach. More importantly, the semi-rigid docking protocol is likely to perform better when no crystal structure information is available. We also present a new parameter for parsing selective versus non-selective GAG–protein systems, which relies on two computational parameters including consistency of binding (i.e., RMSD) and docking score (i.e., GOLD Score). The new semi-rigid protocol in combination with the new computational parameter is expected to be particularly useful in high-throughput screening of GAG sequences for identifying promising druggable targets as well as drug-like Hp/HS sequences.

**Keywords:** heparin/heparan sulfate; molecular docking; glycosaminoglycans; knowledge-based docking



**Citation:** Holmes, S.G.; Desai, U.R. Assessing Genetic Algorithm-Based Docking Protocols for Prediction of Heparin Oligosaccharide Binding Geometries onto Proteins.

*Biomolecules* **2023**, *13*, 1633. <https://doi.org/10.3390/biom13111633>

Academic Editor: Alessandro Paiardini

Received: 12 October 2023

Revised: 6 November 2023

Accepted: 7 November 2023

Published: 9 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

High-throughput molecular docking is an oft-used computational technique in drug discovery, which requires extensive sampling of bound conformations of each molecule in a library and concomitant scoring of the interactions with a putative site of binding on a protein. Popular software packages available for this purpose include AutoDock, MOE, DOCK, GOLD and others, which purport to efficiently sample the conformational space available to both the ligand and the protein in a short time. Unfortunately, this is achievable only if the number of rotatable bonds is not high [1–3], which is typically possible for small, drug-like hydrophobic molecules.

Molecular docking and scoring techniques have also been implemented for biomolecules such as smaller peptides and nucleic acids, which have garnered much interest as therapeutics in recent decades [4–6]. Specialized docking protocols have been designed to address their higher flexibility [7–13]. In fact, these docking protocols have advantageously utilized the vast amount of structural data available for peptides. For example, the ensemble docking of pre-generated conformer libraries based on experimentally determined structures have significantly improved docking efficiency [13,14]. These structural datasets have been leveraged

in developing forcefield parameters for on-the-fly sampling of peptide conformations using physics based-approaches [12,15].

A super-class of arguably more flexible biopolymers—the glycosaminoglycans (GAGs)—is a family of four unique structures including hyaluronic acid, chondroitin/dermatan sulfate, heparin/heparan sulfate (Hp/HS), and keratan sulfate. These biopolymers are made up of repeating disaccharide building blocks in which a glucuronic acid (GlcA), iduronic acid (IdoA) or a galactose residue is connected to either a glucosamine (GlcN) or galactosamine residue through 1→3 or 1→4 inter-glycosidic linkages (Figure S1). Interestingly, except for HA, GAGs can be variably *N*- or *O*- sulfated and *N*-acetylated at various positions by a panel of biosynthetic enzymes in the Golgi, which generates an astounding level of configurational and conformational diversity. For example, nearly 139 billion theoretical topologies are possible for a hexamer of Hp/HS prepared from one of its 72 building blocks assuming occupancy of only two puckers (<sup>1</sup>C<sub>4</sub> and <sup>2</sup>S<sub>O</sub>, Figure S2) for its IdoA residues. By comparison, a hexapeptide or a hexanucleotide may occupy only about 64 million or 4000 topologies, respectively (Table S1).

A major corollary of the topological diversity of GAGs is that these biomolecules recognize and bind to a huge number and diversity of proteins. The latest compilation pegs the GAG interactome to be 3464 strong [16]. Understanding such an interactome necessitates the use of *in silico* approaches, which have already been used in the past to elucidate the GAG-binding potential of proteins. Methods used so far include amino acid primary sequence analysis [17,18], surface electrostatics [19,20], and dynamic molecular dynamics [21,22]; a high-throughput computational approach that rapidly predicts the strength and binding geometry of each GAG sequence in the library of millions is critically needed to identify biologically relevant GAG–protein pairs. Unfortunately, such a computational approach is a pipe-dream at this time. Although dedicated software packages are available, including GlycoTorch Vina [23], Vina-Carb [24], and ClusPro with heparin extension [25,26], basic GAG docking is still challenging and limited.

One major limitation faced by GAG docking approaches is the paucity of highly efficient, and highly reliable, docking protocols. Currently, GAG docking protocols have implemented either “rigid” or “flexible” inter-glycosidic linkages ( $\Phi$  and  $\Psi$ ) to predict the strength and binding geometry of sequences. To a large extent, these approaches have been used to predict the sites of GAG binding, the length and fine structure of the preferred GAG sequence(s) and the native GAG binding pose. A high-throughput protocol put forward in recent times is the combinatorial virtual library screening (CVLS) protocol, which has achieved considerable success in the screening of a library of GAG sequences against different proteins such as antithrombin, spike glycoprotein, heparin cofactor II, transforming growth factor beta, and insulin-like growth factor–1 receptor [27–31]. CVLS employs rigid  $\Phi$  and  $\Psi$  values, which are taken from the literature on related sequences and fixed at the midpoint of the known range. Many other groups, including those of Samsonov, Gandhi, Ricard-Blum, Rusnati, and Pisabarro, have also implemented rigid docking on a number of GAG sequences to much success, including bone morphogenetic protein 2,  $\alpha v\beta 3$  integrin, HIV-1 matrix protein p17, CXCL12 [32–35] and many others.

By the same token, flexible docking has also helped shed light on the GAG recognition of proteins, especially Hp and CS [35–37]. Interesting insights can be gained through flexible docking, as shown in a CCL5 study with a small library of Hp tetrasaccharides that revealed a novel binding mechanism involving the protonated His23, which is populated at pHs less than 7 [37]. Unfortunately, flexible docking generally succeeds for smaller GAG chains [37,38]. Sometimes molecular dynamics is used as a follow up method to “clean-up” top-ranked poses of flexible docking, which significantly restricts the application of this approach under a high-throughput format [36,39,40]. To address this, coarse grain (CG) modeling with full flexibility has been implemented for longer GAG oligosaccharides; yet, even here the CG beads attempt to drastically reduce the flexibility arising from ring substituents [41].

We reasoned that fundamental, comparative studies on rigid and flexible docking protocols are necessary to understand their applicability for the high-throughput screening of a large library of GAG sequences. We also reasoned that an alternate protocol, i.e., semi-rigid docking, be assessed before embarking on designing a robust, high-throughput algorithm. In this work, we report a comparative study of docking 18 Hp/HS oligosaccharides, ranging from di- to decasaccharide, onto their targets using “rigid”, “flexible” and “semi-rigid” protocols. We have analyzed their successes in locating the “native” pose, as observed in the co-crystal structures. We have chosen a structurally diverse range of proteins exhibiting a range of selectivities—from the highly selective (antithrombin) to highly non-selective (thrombin). We find that although all three protocols succeed reasonably well, the rigid and semi-rigid protocols recapitulate crystal structure poses for chains as large as a decasaccharide more often and in a reproducible manner. Our work shows that for an unknown system, the semi-rigid protocol is a better alternative to employ in comparison to the rigid (when the “native”  $\Phi/\Psi$  are not known). Likewise, the semi-rigid docking protocol is much better than the flexible docking protocol, especially for longer GAGs (5→10 mers) and also for smaller sequences (2→4 mers). Finally, we present a computational parameter that could be employed in the virtual high-throughput screening of thousands of sequences for identifying putative drug-like Hp/HS structures.

## 2. Methods

### 2.1. Software

SYBYLX 1.3 (Tripos Associates, St. Louis, MO, USA) was used for molecular visualization, minimization, and protein/ligand preparation. GOLD Version 2020 was used for molecular docking experiments [1].

### 2.2. Structure Selection and Preparation

Co-crystal structures were generally selected based on the reported resolution ( $\leq 2.5$  Å) and containing the entire structure of the sequence used for crystallization (2→10-mer) (Table S4). For 5DNF, 2VRA, 3EVJ, 7B8I, and 1E0O, no oligomer structures were available that met the resolution cut-off and hence the next best co-crystal structure was utilized. Unnecessary subunits, cofactors, and metal ions that did not directly bind to the ligand were removed. The Hp/HS sequence was extracted followed by the removal of water molecules. Hydrogens were added to the protein in SYBYL X1.3 and the protein was minimized with fixed heavy atom coordinates using the Tripos force field for 100,000 iterations followed by a termination gradient of 0.05 kcal/mol. Gasteiger–Hückel charges were introduced and structure minimization was carried out with a non-bonded interaction cut-off of 8 Å and a dielectric constant of 80.

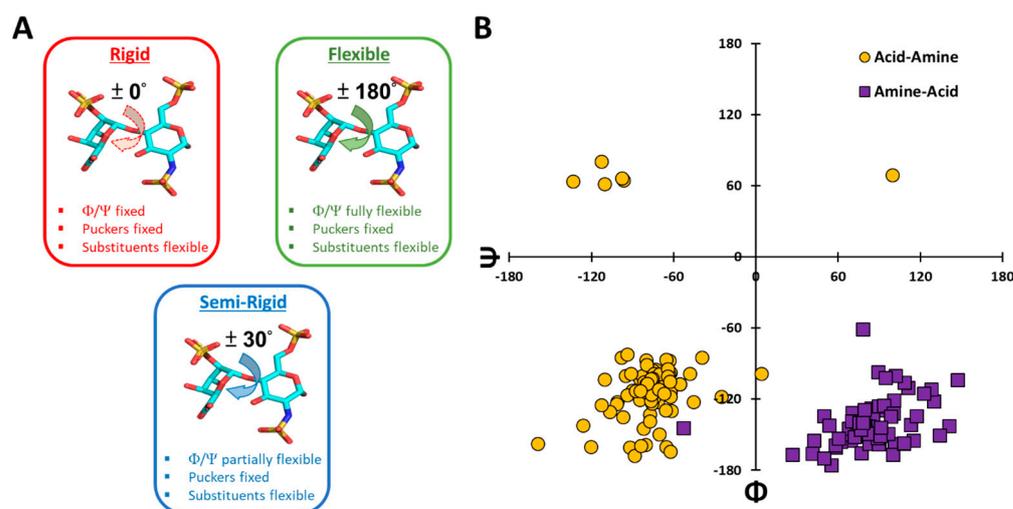
### 2.3. Hp/HS Sequence Preparation

Post extraction of the sequence, structures were manually inspected to ensure correct atom and bond typing. The Tripos force field does not contain parameters for sulfate oxygens, so the carboxylate oxygen (O.CO<sub>2</sub>) atom type was used as a surrogate, as documented in earlier studies for the parameterization of GAGs [30,42]. Hydrogen atoms were added to the ligand in SYBYLX 1.3.

### 2.4. Molecular Docking

The default ligand binding site definition in GOLD was used for all studies in this work. In this definition, all residues that are within 6 Å of the native pose are selected. The protein was kept rigid during the docking, while the Hp/HS ligand was allowed varying degrees of flexibility depending on the protocol being used, except for ring torsions (see Figure 1A). Glycosidic torsions from each crystal structure were used as reference for rigid and semi-rigid dockings. In the rigid docking protocol, the glycosidic torsions were restricted to the initial values, while all other non-terminal single bonds were allowed free rotation. In contrast, flexible docking allows free rotation at glycosidic linkages. For

semi-rigid dockings, the user specifies torsional histograms for  $\Phi$  and  $\Psi$  for each glycosidic linkage, where the center of the bell curve corresponds to the initial value. Although we chose a  $30^\circ$  cutoff in this study, the user can adjust these values if needed. To generate a torsional histogram in GOLD, we refer the reader to the GOLD user guide, but we provide a brief example in the Supplementary Material (Figure S7). For dockings, 100 GA and 300 GA runs were used with 10,000 genetic operations each. Each docking experiment was performed in triplicate and the two poses from each run with the highest GOLD Scores were retained for analysis. A total of six poses per structure were utilized for analysis of docking protocol performance. Pairwise RMSDs and glycosidic torsions were calculated via in-house python scripts. All molecular models were generated in PyMOL.



**Figure 1.** Types of docking protocols used in the literature. (A) Two types of protocols typically used for GAGs in the literature including “rigid” and “flexible” docking. This work reports comparative studies of these two with a semi-rigid docking protocol, which affords better predicting success, especially for longer GAG sequences. Whereas the rigid and flexible docking approaches hold glycosidic torsions ( $\Phi$  and  $\Psi$ ) either completely invariant ( $\pm 0^\circ$ ) or fully flexible ( $\pm 180^\circ$ ), the semi-rigid protocol allows partial flexibility ( $\pm 30^\circ$ ) around the most preferred torsions reported in the literature. (B) A Ramachandran plot depicting  $\Phi$  and  $\Psi$  for all 41 Hp/HS-protein complexes available in the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org), accessed on 1 July 2021).  $\Phi$  and  $\Psi$  are categorized into two groups: acid-amine (UA→GlcN) and amine-acid (GlcN→UA).

### 3. Results

#### 3.1. Basis Underlying the Semi-Rigid Docking Protocol

Although GAGs contain multiple ring substituents ( $-H$ ,  $-\text{OSO}_3^-$ ,  $-\text{NHCOCH}_3$ ,  $-\text{NHSO}_3^-$ ) that are flexible, glycosidic torsions ( $\Phi/\Psi$ ) and ring puckering ( $^1\text{C}_4/{}^2\text{S}_0$ ) are the bane of conformational challenges in docking studies. A typical rigid docking protocol assumes that glycosidic torsions and ring puckers are restricted to their initial values during the search, while imparting flexibility to ring substituents (Figure 1A). In contrast, a fully flexible docking algorithm imposes no constraints on glycosidic torsions and ring substituents. Whereas the rigid docking protocol samples minimal conformational space, the flexible protocol attempts to sample the entire conformational space. Alternatively, the two protocols represent two extreme ends.

It has now been long known that GAG glycosidic torsions in oligosaccharides sample a rather limited range of  $\sim 60^\circ$ . Curating  $\Phi/\Psi$  values from the PDB shows two dominant clusters corresponding to two categories (GlcN→UA and UA→GlcN) with only a few unusual torsions (Figure 1B, Tables S2 and S3). In fact, these torsional preferences were exploited in developing VinaCarb and GlycoTorch Vina (GTV), two computational tools in which low-energy ligand poses are pre-generated using quantum mechanical (QM)

energy wells predicted by glycosidic torsion's specific carbohydrate intrinsic (CHI) energy functions [23,43].

With regard to ring puckers, the flexibility of IdoA is thought to be a key contributor to selective recognition [44–46]. IdoA can sample a number of ring puckers; however,  ${}^1C_4$  and  ${}^2S_0$  have been shown to be the most populated with their dynamic equilibrium determined by the neighboring sulfation pattern [47]. Several authors have explored the role of IdoA puckers in protein recognition [44,48,49]; however, Boittier et al. have used GTV to elucidate the effects of IdoA puckers (especially  ${}^2S_0$ ) on the internal energy of the GAG chain [23]. Interestingly, redocking with recalculated CHIs for seven common GAG linkages using density functional theory resulted in the recapitulation of 5/12 test cases. Unfortunately, GTV accuracy appears to taper off for chains longer than a hexasaccharide.

Given the current state of knowledge and challenges, we reasoned that a simple, rapidly implementable, reasonably comprehensive approach to recapitulate native poses of GAG oligosaccharides would be to rely on torsional probability distributions that are derived from  $\Phi/\Psi$  averages. As shown in Figure 1B, low-energy torsions are found to populate a rather narrow range of  $\pm 30^\circ$  (see also Tables S2 and S3 for all compiled data). Such a probability distribution would afford the flexibility necessary for an optimal fit into the binding pocket, which may be bypassed during a purely rigid docking search. At the same time, the partial flexibility of  $\pm 30^\circ$  would also avoid the massive, and unnecessary, conformational search that accompanies a fully flexible docking search. We hypothesized that such a probabilistic bias of  $\pm 30^\circ$  around the common  $\Phi/\Psi$  average could increase the likelihood of finding the native pose. We call this approach the “semi-rigid” docking (SRD) protocol. However, before implementing such a protocol for high-throughput purposes, key questions need to be addressed, including (i) how useful are rigid and flexible docking approaches? (ii) Can the SRD protocol better for longer GAG chain? and (iii) does it recapitulate the native pose of a GAG–protein complex? This work addresses these and other comparative questions.

### 3.2. Does Rigid Docking Recapitulate the Native Pose?

Although ligand geometries presented in the PDB entries may not necessarily be accurate, the large majority are assumed to be the “native”, most preferred, binding geometries [50]. It is expected that as a ligand–protein complex crystallizes, the “native” pose outcompetes other possible poses because of stereo-electronic restrictions imposed by a well-defined binding site. Unfortunately, most sulfated oligosaccharides bind in a shallow, surface-exposed, flexible, highly water-laden, cationic binding site [20,26,51]. This raises a question as to whether the “selectivity” engineered by crystallization forces represents an intrinsic property of the GAG–protein complex. In fact, the crystal structure of a hedgehog–heparan sulfate complex exhibited dimerization by binding two different, adjacent HS sequences on the same chain with different conformations, despite interacting with the same five residues [52]. Likewise, the prototypic, non-selective, GAG-binder thrombin displays at least two crystal poses for Hp/HS oligomer that differ significantly in torsional and translational profiles (Figure S3). Given this ambiguity, it is useful to assess whether the “native” crystal structure pose is recapitulated upon docking, especially when the GAG is conformationally immobile during docking. As stated above, “rigid” docking holds  $\Phi/\Psi$  and ring puckers invariant during docking and equal to their crystal structure values, while affording flexibility to ring substituents (Figure 1A).

We studied the rigid docking protocol on 18 proteins that had been co-crystallized with an oligosaccharide using GOLD, a genetic algorithm (GA)-based protocol. The Hp/HS oligosaccharides ranged in length from di- to decasaccharide. Their fine structure could be classified into primarily two types including those containing the common repeating sequence and some containing the rare 3-O-sulfated sequence (Table S4). Our docking protocol implemented 100 GA runs with 100,000 operations per run, which has been employed in numerous studies in the literature [53,54] and was found to be very useful for GAGs also [20,30,42]. Experiments were performed in triplicate for each Hp/HS oligosaccharide,

and the top two poses each 100 GA run, i.e., six poses in toto, were collected. Analysis of the collected poses was performed in a quantitative manner through the use of a parameter called RMSD, which stands for root mean square difference, between either the native and predicted poses or between different predicted poses. In the literature, an RMSD of 2.5 Å or less has been deemed as geometric equivalence between two poses [20,55,56]. For most part, the cut-off of 2.5 Å typically works for small molecules and smaller oligosaccharides (<6 mers). It is important to recognize that longer oligosaccharides may not adhere to this arbitrary cut-off. Alternatively, RMSD cut-off may have to be scaled with chain length. Yet, we find that the RMSD cut-off of 2.5 Å worked for several oligosaccharides longer than 6 mers (see below).

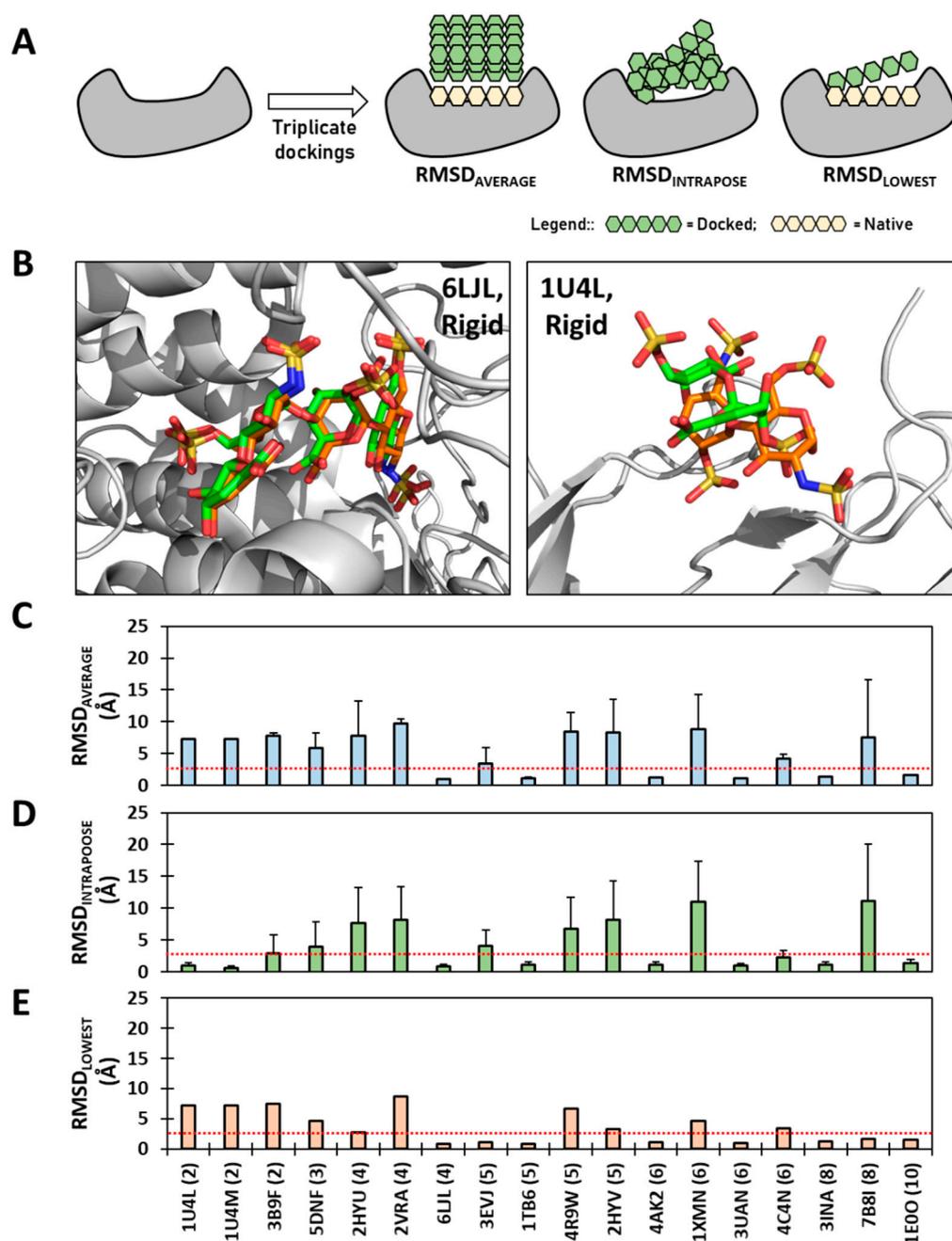
We performed three comparisons of the docked pose(s) with the native pose observed in the co-crystal structures (Figure 2A). (i) We first calculated  $\text{RMSD}_{\text{AVERAGE}}$  as the difference between the native pose and the average of the top six rigid docking poses. (ii) Then,  $\text{RMSD}_{\text{LOWEST}}$  was calculated to quantitatively assess the difference between the native pose and the one docked pose that most closely matches the native. This parameter has been an oft-used method reported in the literature [57–59]. (iii) We finally calculated  $\text{RMSD}_{\text{INTRAPOSE}}$ , which assessed the similarity between the six docked poses. Calculation of these three RMSDs afforded a rather clear quantitative insight into the similarity of recognition, as evident from two representative sequences—a tetrasaccharide 6LJL and a disaccharide 1U4L—which present successful recapitulation ( $\text{RMSD} \leq 2.5 \text{ \AA}$ ) of the native pose and a not-so-successful prediction ( $\text{RMSD} > 2.5 \text{ \AA}$ ), respectively (Figure 2B).

Figure S4 presents the comparative overlays of docked poses using rigid docking with the native pose for all 18 sequences. As evident from visual inspection, and counter to the simplistic expectation, rigid docking more frequently does not recapitulate the native pose found in the PDB co-complex. In fact,  $\text{RMSD}_{\text{AVERAGE}}$  calculation shows that only 6 of 18 HS oligosaccharides (33%) recapitulated the native pose (Figure 2C). Even when only the best pose is compared, as in  $\text{RMSD}_{\text{LOWEST}}$ , recapitulation increases to 44% (Figure 2E). Curiously, none of the di- or trisaccharides recapitulated their native pose, whereas one each of tetra-, octa- and decasaccharide and two of the penta- and hexasaccharides recapitulated the native poses. This is interesting because the di- and trisaccharides are theoretically expected to span much smaller conformational space than the longer oligomers.

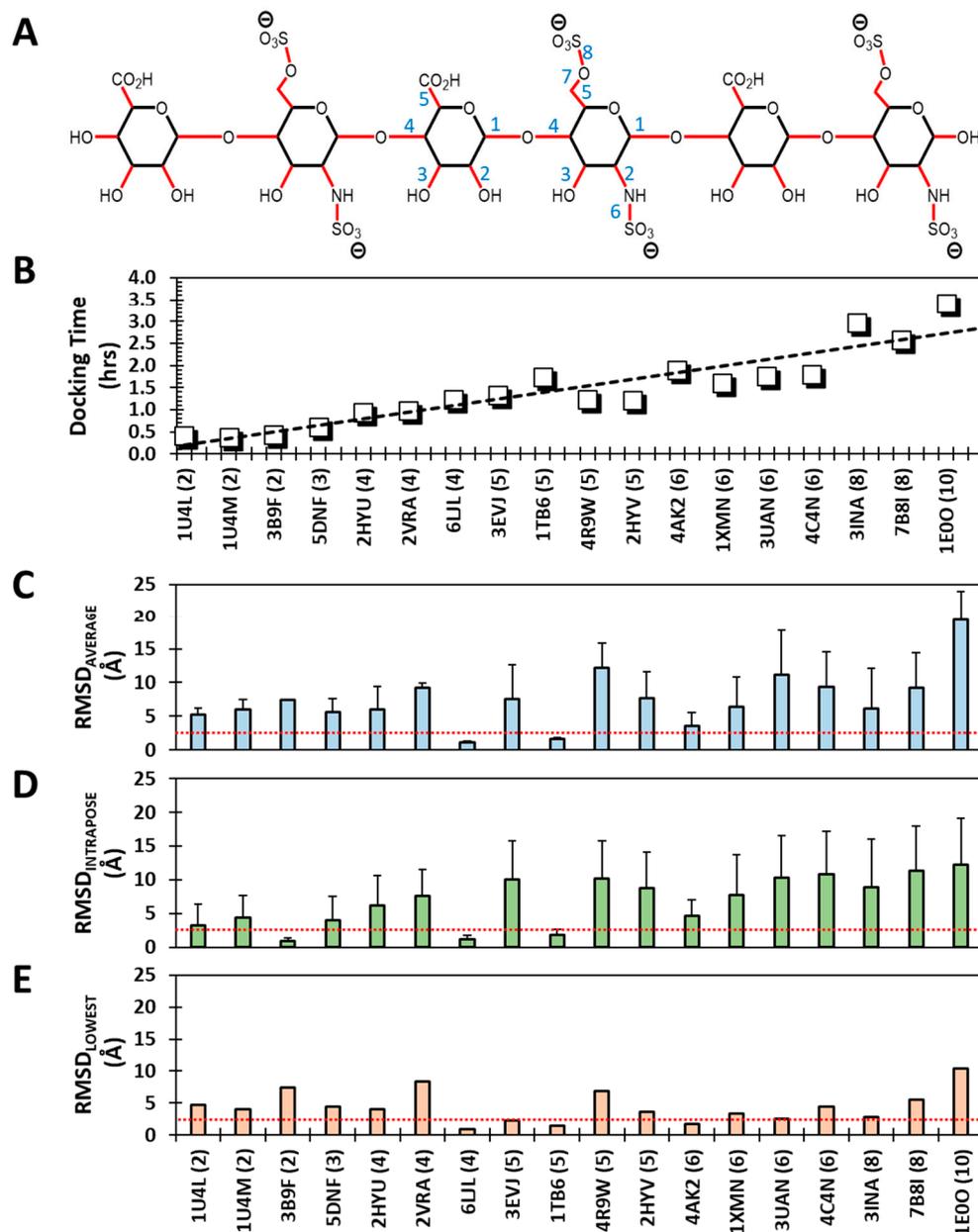
When  $\text{RMSD}_{\text{INTRAPOSE}}$  is calculated, 50% of sequences were found to bind very consistently ( $\text{RMSD} \leq 2.5 \text{ \AA}$ ) (Figure 2D). Of these, 33% recapitulate the native pose (Figure 2C), which implies that three sequences (1U4L(2), 1U4M(2), and 4C4N(6)) bind consistently away from the site of the native ligand. Alternatively, these oligosaccharides converged to a single pose, but this was significantly different from the native pose. We also performed docking experiments using 3-fold-higher GA runs (i.e., 300 GA runs; each with 100,000 genetic operations) to assess whether more sampling helps the recapitulation of the native form; however, this did not yield any significant improvement in recapitulation success (Figure S5). This suggests that our genetic algorithm-based approach is not able to identify the “native” pose in the rigid docking paradigm, especially for 2→4 mers.

### 3.3. How Does Flexible Docking Approach Compare to the Rigid Approach?

Next, we evaluated the impact of imparting full flexibility to all rotatable bonds in an Hp/HS chain. Although GOLD holds pyranose ring puckers invariant from initial assignment, a typical Hp/HS sequence still encompasses at least five and seven rotatable bonds for all the UA and GlcN residues, respectively. This means that flexible docking minimally requires searching conformational space for 12 to 60 rotatable bonds for di- to decasaccharide, respectively (Figure 3A). In fact, the time for a triplicate flexible docking run steadily increases with chain length despite the limitation of 100,000 iterations (Figure 3B). Yet, this may not lead to the effective recapitulation of the native pose; rather, as the chain length increases, one may predict that flexible docking will not effectively recapitulate the native pose.



**Figure 2.** Recapitulation of the native pose using a rigid docking protocol. Each sequence was redocked back into the crystal structure in triplicate using 100 GA runs, each being allowed 100,000 genetic operations. The top two poses from each replicate experiment were selected, compiled and used for analysis. (A) The docking of each Hp/HS oligosaccharide onto its target protein was analyzed by calculating the  $RMSD_{AVERAGE}$ ,  $RMSD_{LOWEST}$  and  $RMSD_{INTRAPOSE}$ , which convey the root mean square difference (RMSD) between the native pose and the average of the top six rigid docking poses, the difference between the native pose and the one docked pose that most closely matches the native, and the intra-pose difference between the six docked poses, respectively. (B) Representative example of a successful recapitulation ( $RMSD_{AVERAGE} \leq 2.5$  Å) of the native pose of an Hp/HS tetrasaccharide (left; 6LJL) and a not-so-successful predication ( $RMSD_{AVERAGE} > 2.5$  Å) of the native pose an Hp/HS disaccharide (right; 1U4L). Native poses in both are shown in green, while docked poses are in orange. (C–E) Three different RMSDs as function IDs of the co-complex structures reported in the PDB. X-axis labels represent the PDB code followed by chain length in brackets. The red dotted line indicates the 2.5 Å cut-off.



**Figure 3.** Recapitulation of the native pose using a flexible docking protocol. Each sequence was redocked back into the crystal structure in triplicate using 100 GA runs, each being allowed 100,000 genetic operations. The top two poses from each replicate experiment were selected, compiled and used for analysis. (A) The flexible docking protocol affords full flexibility to glycosidic bonds and ring substituents (shown in red). A typical HS hexasaccharide encompasses more than 36 rotatable bonds arising from a minimum of 5 and 7 rotatable bonds in UA and GlcN residues (labeled 1→7 in blue). Ring puckers are held invariant from their starting state in the flexible docking protocol. (B) As expected, GOLD dock time for fully flexible docking increased linearly with chain length; although this does not imply that flexible docking yields recapitulation of the native pose (see text for details). (C–E) Three different RMSDs as function IDs of the co-complex structures reported in the PDB. X-axis labels represent the PDB code followed by chain length in brackets. Red dotted line indicates the 2.5 Å cut-off.

Figure 3C–E show RMSD<sub>AVERAGE</sub>, RMSD<sub>INTRAPPOSE</sub> and RMSD<sub>LOWEST</sub> profiles for the 18 proteins at 100 GA runs. RMSDs were within the 2.5 Å threshold for 2, 3 and 5 Hp/HS chains out of 18 when “average”, “intrapose” and “lowest”, respectively, were compared to the native pose. As predicted, these successes are much lower than the corresponding

results for rigid docking (six, eight and nine, respectively). Increasing the number of GA runs to 300 does not yield any significant improvement (Figure S6). Yet, some very interesting results can be gleaned from flexible dockings.

One, the flexible docking of two HS oligosaccharides appears to consistently converge to the native pose. These include a pentasaccharide (1TB6), which is universally recognized as the prototypic high-affinity, high-specificity antithrombin-binding Hp/HS oligosaccharide [60], and a tetrasaccharide cleavage product (6LJL), which is generated by a unique exolytic heparinase [61] with a high level of substrate specificity. Thus, flexible docking appears to work primarily for sequences that exhibit a high level of selectivity.

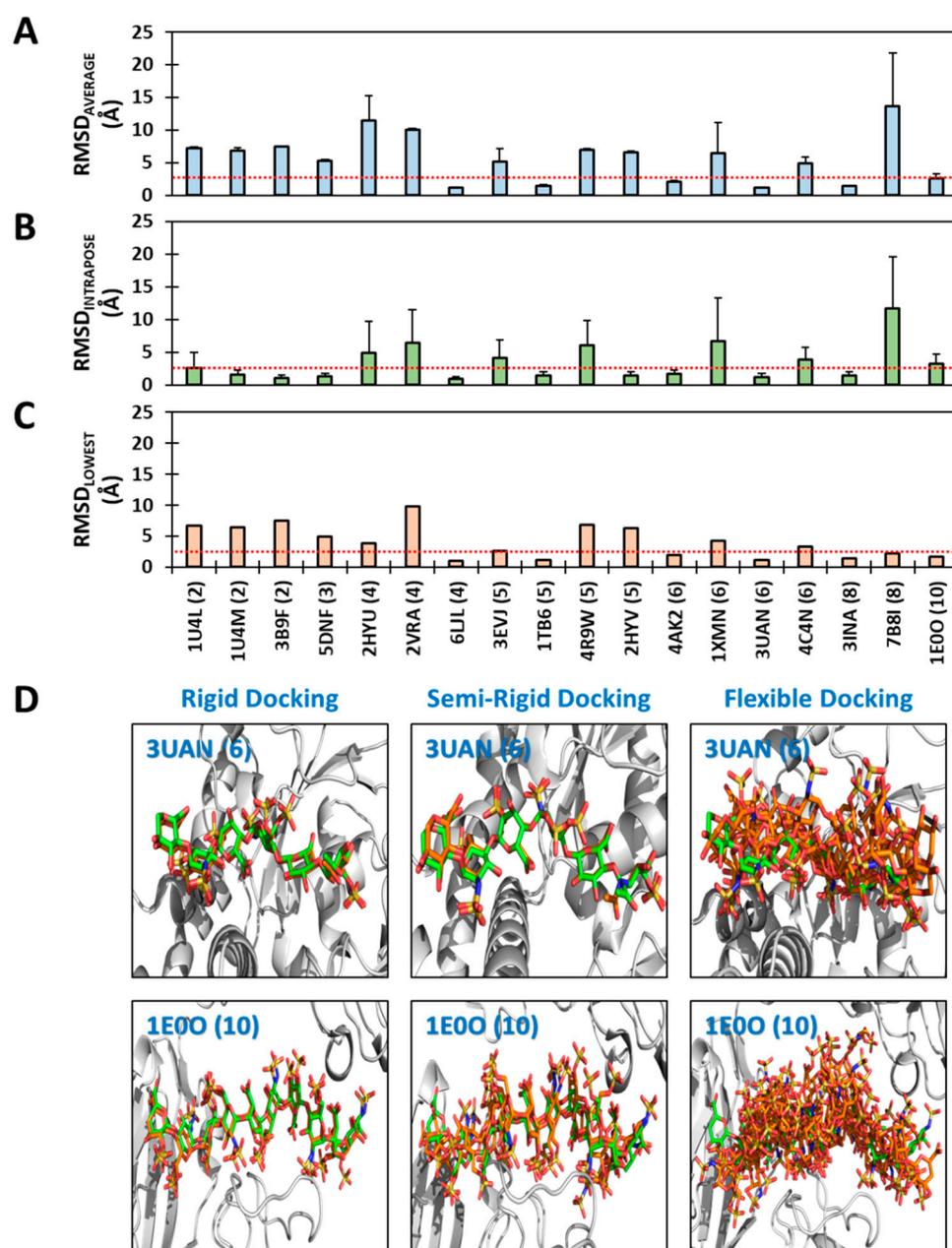
Two, if only the most native-like pose of top six docked poses is taken, as in  $\text{RMSD}_{\text{LOWEST}}$ , the performance improves to 5 out of 18 chains (Figure 3E). Interestingly, these five sequences are tetrasaccharide or longer. Alternatively, none of the di- and trisaccharides recapitulated the native pose, results similar to those in the rigid docking study. In fact, a correspondence between rigid and flexible dockings can be noted for longer HS chains also.

Three, when  $\text{RMSD}_{\text{INTRAPOSE}}$  values are considered, only three HS chains pass the threshold, including 3B9F, 6LJL, and 1TB6. This result is dramatically different from that of the rigid docking protocol, wherein 44% of HS oligosaccharides converged to a single pose. This result reiterates that the flexible docking protocol tends to yield a wider array of significantly different poses, which also turn out to be different from the native pose. Alternatively, the flexible docking of GAGs appears to be inherently beset with a huge conformational space that is difficult to sample comprehensively within a reasonably short time.

### 3.4. Can Semi-Rigid Docking Approach Offer a Better Alternative?

To assess whether a rational, balanced docking approach can offer a better alternative, we studied a semi-rigid docking (SRD) protocol. As described in the section on rationale above, this protocol relies on the well-established understanding that glycosidic torsions  $\Phi$  and  $\Psi$  prefer a rather narrow range of  $\pm 30^\circ$ , irrespective of the local structural heterogeneity (Figure 1B, Tables S2 and S3). More specifically, it is usually  $\Psi$ , rather than  $\Phi$ , that tends to vary more [23,62,63]. Further, two fairly defined  $\Phi$  and  $\Psi$  minima are generally observed for the more flexible UA $\rightarrow$ GlcN torsions ( $\sim -80^\circ / -100^\circ$  and  $\sim -80^\circ / 65^\circ$ ), whereas only one  $\Phi$  and  $\Psi$  minimum is typically populated for GlcN $\rightarrow$ UA torsions ( $\sim 80^\circ / -145^\circ$ ) [62] (Figure 1B). Thus, a better docking approach would be to pre-generate torsional probability distributions around the well-known  $\Phi$  and  $\Psi$ , then harness the power of GA to enhance the probability of finding the native pose. Although the current approach utilizes  $\Phi$  and  $\Psi$  torsions derived from crystallographic and/or NMR studies, in principle, other computational calculations such as DFT, QM, and/or MM [64–67] may also be used. Briefly, our SRD protocol utilized a torsional probability distribution function (Figure S7) around pre-chosen  $\Phi$  and  $\Psi$  torsions corresponding to the typical minima observed in solution. GOLD docking and analysis was performed as described for the rigid and flexible docking methods above.

Figure 4 presents the  $\text{RMSD}_{\text{AVERAGE}}$ ,  $\text{RMSD}_{\text{LOWEST}}$  and  $\text{RMSD}_{\text{INTRAPOSE}}$  analyses following SRD re-dockings using 100 GA runs. This protocol demonstrated accuracies similar to that of the rigid protocol, with rates of 33%, 44%, and 55%, respectively, which did not change even at 300 GA runs (see Figure S18). Interestingly, the SRD was also unable to recapitulate the native pose for smaller chains, i.e., all di-/tri- and most tetra- saccharides (Figure 4A). Yet, SRD was much better than the flexible protocol, especially for sequences longer than pentasaccharide. In fact, for longer chains ( $\geq 5$  mers),  $\text{RMSD}_{\text{AVERAGE}}$  exhibits a success rate of 45% for the SRD protocol in comparison to 9% for the flexible approach (Figures 3C and 4A). More importantly, SRD was able to predict the crystal structure poses of longer sequences very well, as exemplified by 3UAN and 1E0O (Figure 4D).

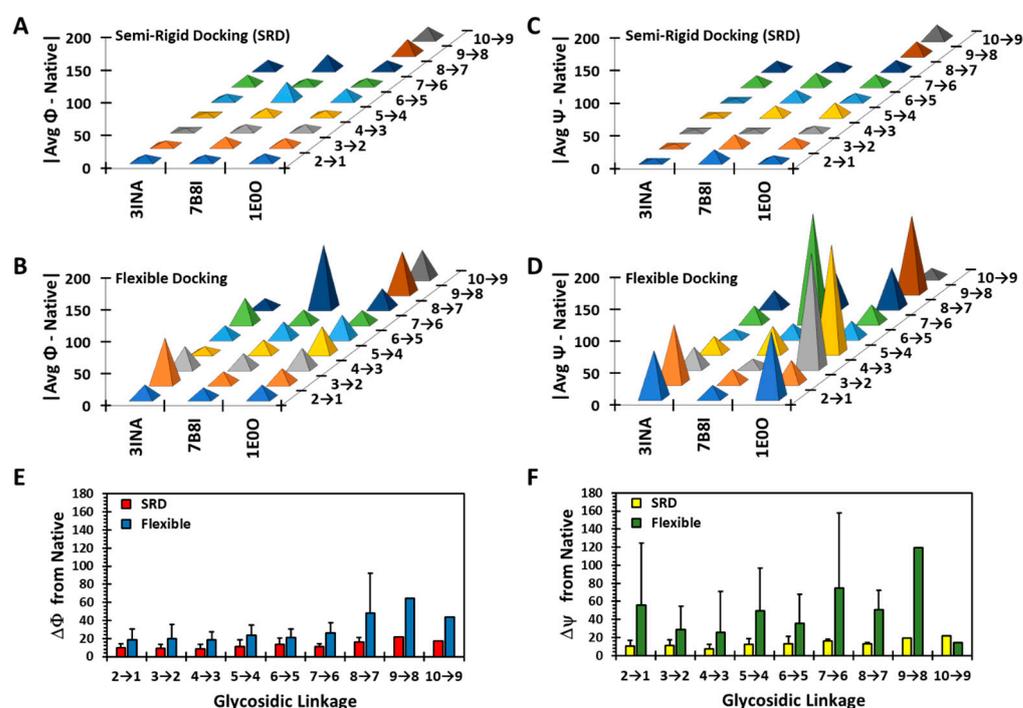


**Figure 4.** Recapitulation of the native pose using a semi-rigid docking (SRD) protocol. As for rigid and flexible protocols, each sequence was redocked back into the crystal structure in triplicate using 100 GA runs, each being allowed 100,000 genetic operations. (A–C) Three different RMSDs as function IDs of the co-complex structures reported in the PDB. X-axis labels represent the PDB code followed by chain length in brackets. The red dotted line indicates the 2.5 Å cut-off. (D) Successful recapitulation of native poses of 3UAN and 1E0O by rigid (left) and semi-rigid (middle) docking protocols but not by the flexible docking protocol (right). Docked poses (shown in orange) are superimposed on native poses (green) for the two sequences. The protein ribbon is shown in light grey.

The RMSD<sub>INTRAPOSE</sub> analysis of the SRD results reveals some interesting insights. Four of six smaller chains (2→4-mers) demonstrate highly consistent recognition, albeit different from the native pose (Figure 4B). This proportion is nearly two-fold higher than that observed for rigid docking (71% vs. 43%), suggesting that limited flexibility is important for convergence. In fact, the SRD approach is more effective in identifying proteins that exhibit significant levels of consistent recognition in comparison to rigid and flexible protocols. For

example, trisaccharide 5DNF displayed  $\text{RMSD}_{\text{INTRAPOSE}}$  of 1.28 Å for SRD in comparison to 3.95 Å and 4.02 Å for rigid and flexible protocols, respectively. Similar results were observed for the 3B9F and 2HYV sequences.

To gain more insight into variances from consistent recognition, we calculated the deviation of each torsion from the native pose following SRD and flexible docking. Figure 5 presents the results for two octa- and one decasaccharide, whereas Figure S19 presents the results for all 18 chains. Although both  $\varphi$  and  $\psi$  deviated more for the flexible protocol in comparison to the semi-rigid,  $\psi$  exhibited huge deviations (Figure 5C,D), which appears to be the foundational reason for the lack of recapitulation of the native pose. When the differences are averaged across all glycosidic linkages, the flexible protocol consistently predicts deviations in excess of 20° (Figure 5E,F). Interestingly, the deviations in  $\varphi$  increase with the length of the chain, while those for  $\psi$  remain high essentially independently of chain length. These results convey the importance of affording only limited flexibility, if any, to both glycosidic torsions, especially  $\psi$ , during docking operations for better success in native pose predictions.

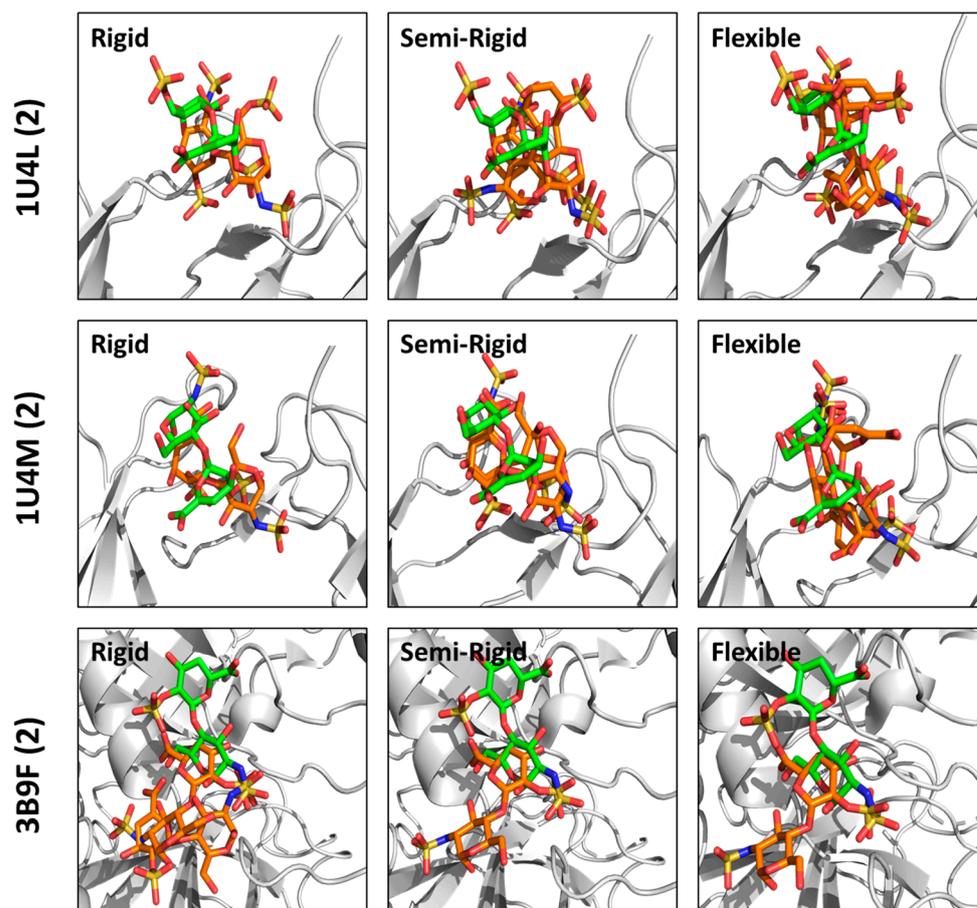


**Figure 5.** Variation in observed  $\Phi/\Psi$  from the native pose following semi-rigid and flexible dockings. Shown are  $\Delta\varphi$  and  $\Delta\psi$ , the differences in  $\varphi$  (A,C,E) and  $\psi$  (B,D,F) between the native pose and the average of the docked poses obtained following semi-rigid (A,C) and flexible (B,D) dockings, respectively, as a function of the co-complex structure and glycosidic bonds (2→1, 3→2, etc., where 2→1 refers to the glycosidic bond between the reducing end residue #1 and the penultimate residue #2). Although the difference ( $\Delta\varphi$  and  $\Delta\psi$ ) could be either negative or positive, only the magnitude is shown (i.e., mod of  $\Delta\varphi$  and  $\Delta\psi$ ). (E,F) show the average  $\Delta\varphi$  and  $\Delta\psi$ , respectively, across all 18 sequences from di- to decasaccharide for SRD and flexible docking protocols. See text for details.

### 3.5. The Enigma of Disaccharides Finding the Native Pose?

A priori, docking a disaccharide onto a pre-determined site of binding should be the easiest and most likely to succeed because of its small size, smaller conformational search space, etc. The literature reports many studies on Hp/HS disaccharides binding to different proteins [68–71] Yet, none of these studies appear to have studied docking consistency with the co-crystal structure. In this connection, we were very surprised with the results for disaccharides observed in this study. None of our three protocols succeeded to any extent. In fact, the  $\text{RMSD}_{\text{AVERAGE}}$  across the library of sequences ranged from 5 to

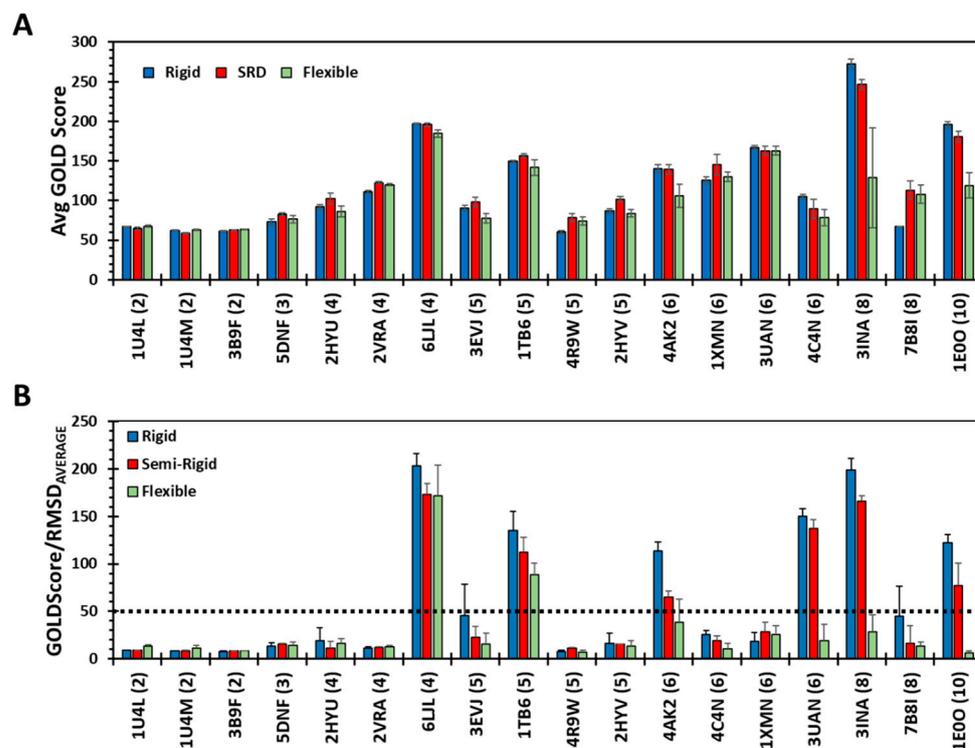
8 Å, which was several-fold higher than the threshold ( $\leq 2.5$  Å) set for the assessment of consistency. Figure 6 shows a comparison of the poses observed following rigid, SRD and flexible docking to the native, co-crystal structure pose for all three disaccharides studied (see Figures S20–S22 for other 15 sequences). All three disaccharides, and especially the 3B9F sequence, prefer to bind in a pocket adjacent to the native binding pocket. Thus, our results raise two possibilities: (i) either crystallization induces smaller sulfated glycans into non-native poses, or (ii) the GA-based docking algorithm fails miserably because the in silico affinity of smaller glycans is not high.



**Figure 6.** Comparison of docked poses with the native pose of disaccharides. The disaccharide sequence from 1U4L, 1U4M and 3B9F was docked onto the protein in triplicate using 100 GA runs (100,000 genetic operations) using either rigid, semi-rigid or flexible docking protocols. The top two poses from each replicate experiment were selected, compiled and used for visualization. The native pose is shown in green. Docked poses are in orange.

To assess the two scenarios, we first reviewed the  $\text{RMSD}_{\text{INTRAPOSE}}$  calculated for the three disaccharides. For rigid and SRD protocols, these RMSDs were noted to be within the set threshold ( $\leq 2.5$  Å), suggesting that the three disaccharides appear to recognize their target proteins consistently (Figures 2D and 4B), albeit in non-native sites. Even for the flexible docking protocol,  $\text{RMSD}_{\text{INTRAPOSE}}$  was found to be very consistent for one of the three disaccharides (Figure 3D). Over the past decade or so, we have demonstrated that a high level of docking consistency, i.e.,  $\text{RMSD} \leq 2.5$  Å, correlates well with a high level of selectivity of protein recognition [20,30,42,49]. Thus, it is very likely that the three disaccharides demonstrate non-native binding poses in the co-crystal form, while in solution they are likely to engage an altered site. Another support for this hypothesis arises from the GOLD Score, a measure of in silico affinity (Figure 7A). The three disaccharides demonstrate a reasonably high GOLD Score, which implies that

the second possibility presented above is unlikely to be true. Thus, our studies raise an alert for both biologists and molecular modelers to be particularly careful in interpreting experiments on smaller sulfated glycans. Rather, our studies point to the importance of simultaneously performing both crystallography and computational experiments, especially on smaller sulfated glycans.



**Figure 7.** Identifying high-affinity, high-selectivity GAG sequences. (A) Average GOLD Scores calculated for six docked poses following rigid, semi-rigid and flexible docking of the 18 Hp/HS sequences onto their target proteins. GOLD Scores reported here were calculated for the protocols with 100 GA runs, each with 100,000 operations. X-axis labels represent the PDB code followed by chain length in brackets. Errors show the standard deviation of scores observed in triplicate docking experiments. (B) A plot of the ratio of the average GOLD Score to  $RMSD_{AVERAGE}$  for rigid, semi-rigid and flexible docking of the 18 Hp/HS sequences. The black dotted line shows an arbitrary cut-off (Ratio  $\geq 50$ ) that can be used to identify high-affinity, high-selectivity sequences. From the 18 sequences studied here, only one tetra- (6LJL), one penta- (1TB6), two hexa- (4AK2 and 3UAN), one octa- (3INA) and one deca-saccharide (1E0O) are predicted to pass the threshold.

### 3.6. A Parameter for Identifying Putative Drug-Like GAG Sequences

A key question stymieing efforts on discovering drug-like GAGs is identifying protein targets that bind sulfated sequences in a highly selective manner. As stated above, we have previously used RMSD as a measure of selectivity for parsing a library of thousands of sequences into selective and non-selective ones [20,27,42]. Do any of these exhibit drug-like characteristics? This work affords a unique opportunity to parse the 18 Hp/HS sequences into putative drug-like sequences. A re-review of results (Figures 2C and 4A) for rigid and SRD docking protocols reveals that one tetra- (6LJL), one penta- (1TB6), two hexa- (4AK2 & 3UAN), one octa- (3INA) and one deca-saccharide (1E0O) exhibit an  $RMSD_{AVERAGE}$  less than the pre-set threshold for selectivity ( $\leq 2.5$  Å). On the other hand, the flexible docking protocol predicts only two high-selectivity sequences (6LJL & 1TB6) (Figure 3C). Yet, for drug-like properties, high docking scores are critical because these serve as surrogates for solution affinity. Unfortunately, the non-selective recognition of GAGs can also yield high docking scores, as shown by the disaccharides (Figure 6A). We reasoned that the ratio of GOLD Score to RMSD would emphasize both high-affinity and

high-selectivity, and thereby better identify drug-like sequences from the rest. This new parameter (GOLDScore/RMSD) was found to vary from a low of  $\sim 7$  to a high of  $\sim 200$  across the three docking protocols (Figure 7B). Interestingly, the most selective six sequences identified by RMSD analysis also stood out above the others, especially for rigid and SRD protocols. In a quantitative manner, the GOLDScore/RMSD parameter appears to have a clear threshold of 50 for segregating selective versus non-selective systems (dotted line in Figure 7B). Such a quantitative threshold, if supported in future experiments, is likely to provide excellent insight into the GAG recognition of proteins.

#### 4. Discussion

Studying GAG binding to proteins has remained challenging for multiple reasons, of which the difficult synthesis of a library of homogeneous sequences [72,73] and the lack of rigorous computational tools for high-throughput studies [41] are the primary barriers. In the latter case, the two major hurdles have been the large number of rotatable bonds within a relatively small structural frame and the shallow, surface-exposed, cationic binding sites on proteins. A good number of efforts have been directed to resolve these challenges including hybrid fragment-based/coarse grain docking [41], QM restraints on glycosidic torsions [23,43], and molecular dynamics methods [21]. In this context, our current work attempts to show that a knowledge-driven algorithm based on pre-generated low-energy glycosidic torsions offers an attractive alternative for most chains, especially those that are longer. Following a rigorous comparative study on a structurally diverse group of proteins exhibiting diversity of GAG recognition selectivities, we conclude the rigid and semi-rigid protocols recapitulate crystal structure poses for longer chains (5 $\rightarrow$ 10 mers) more often. Between the two protocols, differences observed for smaller sequences convey the special value of the SRD approach over the rigid docking approach. More importantly, for systems lacking co-crystal structure information (i.e., “native”  $\Phi/\Psi$  not known), the SRD protocol is a safer alternative to employ over the rigid protocol. Likewise, the SRD is a much better approach than flexible docking for all oligosaccharide sequences, but especially for longer GAGs (5 $\rightarrow$ 10 mers).

Our work has revealed some interesting results and insights into the different docking protocols and GAG recognition. One striking finding was that smaller oligosaccharides appear to bind to sites other than the crystal structure-determined binding sites, irrespective of the protocol used. Most probably, this arises from crystallization artifacts and will have to be followed up in rigorous computational and solution-based validation experiments.

A key observation of this work was that several co-crystal poses simply could not be recapitulated in any of the three docking approaches. These include RANTES, thrombin, platelet factor 4 (PF4), annexin A2, robo, and hedgehog. While it is difficult to ascribe the exact underpinnings of this result, several possibilities can be postulated. First, the algorithm used in all three protocols, i.e., the GA-based approach with an artificially defined limit of termination, is not sufficient to sample the entire conformational space of longer chain oligosaccharides, i.e., hexasaccharides. Second, the energy terms of the docking program (GOLD) may be better suited for certain types of interactions, e.g., hydrophobic, and not for others, e.g., ionic. Third, our algorithm excluded the effects of hydration and protein flexibility to cut down on computational costs, despite their well-established importance in ligand binding [74–76]. Fourth, we used Gasteiger–Hückel partial charges and a particular force field (Tripos) to parameterize each sequence. These charges and forcefields are estimations at best and may carry inaccuracies in energy calculations of GAG–protein systems, as any other parameterization protocols would carry. It is likely that GAG-specific parameters would be more suitable. Finally, the nature of Hp/HS binding to these proteins is likely to be predominantly non-selective, which could result in many equivalent, but non-identical, binding poses [46]. In fact, RANTES [37,77,78], thrombin [79–81], PF4 [82–85], annexin A2 [86–88], robo [89–92], and hedgehog [52,93,94] appear to not bind to a specific Hp/HS motif (Table S3). In fact, RANTES and thrombin have been shown to bind GAGs other than Hp/HS [95–97]. Of all the above explanations, the final explanation presents a significant

new insight into the GAG recognition of proteins with possible biological consequences. Yet, much work is needed on all the above explanations to gain the confidence regarding the new insight on the GAG recognition of proteins.

To our knowledge, this work on the SRD protocol is the first study of its kind for GAGs, although it has been used in the fields of proteins/peptides, nucleic acid docking and the de novo design of protein binders [8,98–100]. Glycosidic  $\Phi/\Psi$  minima can be availed from NMR, crystallographic, and/or QM/MM methods. In this work, we have primarily relied on crystallographic reports from the protein data bank (PDB), which are not very structurally diverse. Computational experiments have documented variations in preferred  $\Phi/\Psi$  for some 3-*O*-sulfated sequences, i.e., compact geometries [62]. It would be important to study and validate the applicability of the SRD protocol for such alternative geometries too.

In this study, we selected GOLD as the docking platform and utilized Gasteiger–Hückel charges, because together they have been successfully used in the past in identifying selective sequences, which were validated through biophysical experiments [28,30,101]. Yet, the performance of GOLD has not been fully investigated in a comparative manner for GAGs, as done for other docking programs [102,103]. Thus, a comparative study of different parameters and docking programs in the application of the SRD protocol for the prediction of GAG binding geometries should be performed in the future. Likewise, it would also be important to test the applicability of the SRD protocol using GAG-specific parameters, including partial charges, which are absent in the current protocol [26,104]. We also emphasize that protein flexibility is currently neglected in the SRD protocol, and is likely to be important in GAG binding. Protein flexibility is often neglected in the virtual screening of GAGs to make the protocol computationally tractable. However, a two-step approach could be envisaged wherein the protein is held rigid in the first step to identify high-scoring sequences, which are then studied in the second step with binding site flexibility.

A specific point related to the use of other docking platforms is whether multiple top geometries are accessible in the tool (and not necessarily only the topmost geometry). Here, each docking platform inherently calculates such geometries, which our work attempts to convey are important to analyze in a logical manner for specificity analysis. Thus, the principles enunciated here should be applicable irrespective of the docking platform used for the prediction of GAG, peptide, or drug complexes with proteins. Finally, the fundamental work continuing in the direction of developing better force-field parameters for GAGs, e.g., GLYCAM and others [46,47] will help build a better SRD protocol in the future.

This work presents a very useful parameter—GOLDScore/RMSD—that can be used to parse selective versus non-selective GAG–protein systems. In this work, we utilized  $\text{RMSD}_{\text{AVERAGE}}$  to identify some highly selective GAG–protein systems. Yet, it is important to speculate whether  $\text{RMSD}_{\text{AVERAGE}}$  or  $\text{RMSD}_{\text{INTRAPOSE}}$  should be used. We used  $\text{RMSD}_{\text{AVERAGE}}$  because co-crystal structures were available in the PDB. However, a large majority of GAG–protein complexes have not been crystallized. Thus, computationally derived  $\text{RMSD}_{\text{INTRAPOSE}}$  will be the only option for exploratory studies. While it may appear that the lower values of  $\text{RMSD}_{\text{INTRAPOSE}}$  may help identify more selective sequences, it is important to note that GOLD Score is also an important determinant in recognition and has to be reasonably high for the suggested threshold of 50. However, more work is needed to assess the validity of this parameter and the suggested threshold. Yet, this simple parameter is likely to find major value in virtual high-throughput screening studies for the rapid identification of putative drug-like sequences against targets of interest.

## 5. Conclusions

This work presents foundational experimentation on developing a protocol for use in the high-throughput screening of Hp/HS oligosaccharides for protein recognition. Protein binding sites generally accommodate Hp/HS chains that are 4- to 8 mers in

length [28,101,105,106]. The majority of proteins have not been crystallized with GAG partners. Thus, “native”  $\Phi/\Psi$  remain unknown. Thus, for de novo applications, SRD seems to be more suitable. In this connection, the ratio parameter based on GOLD Score and RMSD values, which could be easily implemented in the virtual high-throughput screening of thousands of sequences, would be particularly useful in identifying putative drug-like Hp/HS structures.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biom13111633/s1>.

**Author Contributions:** S.G.H. performed experiments, analyzed data, prepared figures, wrote the first draft of manuscript and implemented revisions. U.R.D. analyzed data, prepared figures, revised and finalized the manuscript and secured funding. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by NIH grants P01 HL 107152, K12 HL 141954, U01 CA 241951 and P01 HL 151333 to URD.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All raw data for 100 and 300 GA runs including configuration files, torsional distribution files used in SRD, output files (i.e., docked poses, rankings, etc.), native poses of ligands, ligand files and protein files can be requested from authors. These files have also been deposited with the publisher in two compressed files named 100\_GA\_raw.zip and 300\_GA\_raw.zip and are available online for free download.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

combinatorial virtual library screening, CVLS; glycosaminoglycans, GAGs; genetic algorithm, GA; glucosamine, GlcN; glucuronic acid, GlcA; heparin, Hp; heparan sulfate, HS; iduronic acid, IdoA; root mean square difference, RMSD; semi-rigid docking, SRD; uronic acid, UA.

## References

1. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748. [[CrossRef](#)] [[PubMed](#)]
2. Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288. [[CrossRef](#)] [[PubMed](#)]
3. Morris, G.M.; Ruth, H.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [[CrossRef](#)] [[PubMed](#)]
4. Fosgerau, K.; Hoffmann, T. Peptide therapeutics: Current status and future directions. *Drug Discov. Today* **2015**, *20*, 122–128. [[CrossRef](#)] [[PubMed](#)]
5. Kazmirchuk, T.D.D.; Bradbury-Jost, C.; Withey, T.A.; Gessese, T.; Azad, T.; Samanfar, B.; Dehne, F.; Golshani, A. Peptides of a feather: How computation is taking peptide therapeutics under its wing. *Genes* **2023**, *14*, 1194. [[CrossRef](#)]
6. Sherman, M.; Contreras, L. Computational approaches in design of nucleic acid-based therapeutics. *Curr. Opin. Biotechnol.* **2018**, *53*, 232–239. [[CrossRef](#)]
7. Dagliyan, O.; Proctor, E.A.; D’Auria, K.M.; Ding, F.; Dokholyan, N.V. Structural and dynamic determinants of protein-peptide recognition. *Structure* **2011**, *19*, 1837–1845. [[CrossRef](#)]
8. Huang, S.Y.; Zou, X. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res.* **2014**, *42*, e55. [[CrossRef](#)]
9. Sun, L.; Fu, T.; Zhao, D.; Fan, H.; Zhong, S. Divide-and-link peptide docking: A fragment-based peptide docking protocol. *Phys. Chem. Chem. Phys.* **2021**, *23*, 22647–22660. [[CrossRef](#)]
10. Tuszyńska, I.; Magnus, M.; Jonak, K.; Dawson, W.; Bujnicki, J.M. NPdock: A web server for protein-nucleic acid docking. *Nucleic Acids Res.* **2015**, *43*, W425–W430. [[CrossRef](#)]
11. Yan, Y.; Zhang, D.; Huang, S.Y. Efficient conformational ensemble generation of protein-bound peptides. *J. Cheminform.* **2017**, *9*, 59. [[CrossRef](#)] [[PubMed](#)]

12. Zhang, Y.; Sanner, M.F. AutoDock CrankPep: Combining folding and docking to predict protein-peptide complexes. *Bioinformatics* **2019**, *35*, 5121–5127. [[CrossRef](#)] [[PubMed](#)]
13. Zhou, P.; Jin, B.; Li, H.; Huang, S.Y. HPEPDOCK: A web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res.* **2018**, *46*, W443–W450. [[CrossRef](#)] [[PubMed](#)]
14. Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S.Y. HDOCK: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **2017**, *45*, W365–W373. [[CrossRef](#)] [[PubMed](#)]
15. Hetényi, C.; van der Spoel, D. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **2009**, *11*, 1729–1737. [[CrossRef](#)]
16. Vallet, S.D.; Berthollier, C.; Ricard-Blum, S. The glycosaminoglycan interactome 2.0. *Am. J. Physiol.-Cell Physiol.* **2022**, *322*, C1271–C1278. [[CrossRef](#)]
17. Cardin, A.D.; Weintraub, H.J. Molecular modeling of protein-glycosaminoglycan interactions. *Arterioscler. Off. J. Am. Heart Assoc. Inc.* **1989**, *9*, 21–32. [[CrossRef](#)]
18. Rudd, T.R.; Preston, M.D.; Yates, E.A. The nature of the conserved basic amino acid sequences found among 437 heparin binding proteins determined by network analysis. *Mol. BioSyst.* **2017**, *13*, 852–865. [[CrossRef](#)]
19. Kogut, M.M.; Marcisz, M.; Samsonov, S.A. Modeling glycosaminoglycan–protein complexes. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102332. [[CrossRef](#)]
20. Sankaranarayanan, N.V.; Nagarajan, B.; Desai, U.R. So you think computational approaches to understanding glycosaminoglycan–protein interactions are too dry and too rigid? Think again! *Curr. Opin. Struct. Biol.* **2018**, *50*, 91–100. [[CrossRef](#)]
21. Marcisz, M.; Gaardlæs, M.; Bojarski, K.K.; Siebenmorgen, T.; Zacharias, M.; Samsonov, S.A. Explicit solvent repulsive scaling replica exchange molecular dynamics (RS-REMD) in molecular modeling of protein-glycosaminoglycan complexes. *J. Comput. Chem.* **2022**, *43*, 1633–1640. [[CrossRef](#)] [[PubMed](#)]
22. Samsonov, S.A.; Gehrcke, J.-P.; Pisabarro, M.T. Flexibility and explicit solvent in molecular-dynamics-based docking of protein-glycosaminoglycan systems. *J. Chem. Inf. Model.* **2014**, *54*, 582–592. [[CrossRef](#)] [[PubMed](#)]
23. Boittier, E.D.; Burns, J.M.; Gandhi, N.S.; Ferro, V. GlycoTorch Vina: Docking Designed and Tested for Glycosaminoglycans. *J. Chem. Inf. Model.* **2020**, *60*, 6328–6343. [[CrossRef](#)] [[PubMed](#)]
24. Nivedha, A.K.; Thieker, D.F.; Makeneni, S.; Hu, H.; Woods, R.J. Vina-carb: Improving glycosidic angles during carbohydrate docking. *J. Chem. Theory Comput.* **2016**, *12*, 892–901. [[CrossRef](#)] [[PubMed](#)]
25. Kozakov, D.; Hall, D.R.; Xia, B.; Porter, K.A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein–protein docking. *Nat. Protoc.* **2017**, *12*, 255–278. [[CrossRef](#)]
26. Mottarella, S.E.; Beglov, D.; Beglova, N.; Nugent, M.A.; Kozakov, D.; Vajda, S. Docking server for the identification of heparin binding sites on proteins. *J. Chem. Inf. Model.* **2014**, *54*, 2068–2078. [[CrossRef](#)]
27. Boothello, R.S.; Sankaranarayanan, N.V.; Sistla, J.C.; Nagarajan, B.; Sharon, C.; Chittum, J.E.; Niyaz, R.Y.; Roy, S.; Nandi, A.; O’Hara, C.P.; et al. Glycan Modulation of Insulin-like Growth Factor-1 Receptor. *Angew. Chem.-Int. Ed.* **2022**, *61*, e202211320. [[CrossRef](#)]
28. Chittum, J.E.; Sankaranarayanan, N.V.; O’Hara, C.P.; Desai, U.R. On the selectivity of heparan sulfate recognition by SARS-CoV-2 spike glycoprotein. *ACS Med. Chem. Lett.* **2021**, *12*, 1710–1717. [[CrossRef](#)]
29. Raghuraman, A.; Mosier, P.D.; Desai, U.R. Understanding dermatan sulfate-heparin cofactor II interaction through virtual library screening. *ACS Med. Chem. Lett.* **2010**, *1*, 281–285. [[CrossRef](#)]
30. Sankaranarayanan, N.V.; Desai, U.R. Toward a robust computational screening strategy for identifying glycosaminoglycan sequences that display high specificity for target proteins. *Glycobiology* **2014**, *24*, 1323–1333. [[CrossRef](#)]
31. Sankaranarayanan, N.V.; Nagarajan, B.; Desai, U.R. Combinatorial virtual library screening study of transforming growth factor- $\beta$ 2–chondroitin sulfate system. *Int. J. Mol. Sci.* **2021**, *22*, 7542. [[CrossRef](#)] [[PubMed](#)]
32. Ballut, L.; Sapay, N.; Chautard, É.; Imberty, A.; Ricard-Blum, S. Mapping of heparin/heparan sulfate binding sites on  $\alpha$ v $\beta$ 3 integrin by molecular docking. *J. Mol. Recognit.* **2013**, *26*, 76–85. [[CrossRef](#)] [[PubMed](#)]
33. Bugatti, A.; Giagulli, C.; Urbinati, C.; Caccuris, F.; Chiodelli, P.; Oreste, P.; Fiorentini, S.; Orro, A.; Milanese, L.; D’Urso, P.; et al. Molecular interaction studies of HIV-1 matrix protein p17 and heparin: Identification of the heparin-binding motif of p17 as a target for the development of multitarget antagonists. *J. Biol. Chem.* **2013**, *288*, 1150–1161. [[CrossRef](#)] [[PubMed](#)]
34. Gandhi, N.S.; Mancera, R.L. Prediction of heparin binding sites in bone morphogenetic proteins (BMPs). *Biochim. Biophys. Acta-Proteins Proteom.* **2012**, *1824*, 1374–1381. [[CrossRef](#)] [[PubMed](#)]
35. Panitz, N.; Theisgen, S.; Samsonov, S.A.; Gehrcke, J.P.; Baumann, L.; Bellmann-Sickert, K.; Köhling, S.; Teresa Pisabarro, M.; Rademann, J.; Huster, D.; et al. The structural investigation of glycosaminoglycan binding to CXCL12 displays distinct interaction sites. *Glycobiology* **2016**, *26*, 1209–1221. [[CrossRef](#)]
36. Pichert, A.; Samsonov, S.A.; Theisgen, S.; Thomas, L.; Baumann, L.; Schiller, J.; Beck-Sickinger, A.G.; Huster, D.; Pisabarro, M.T. Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. *Glycobiology* **2012**, *22*, 134–145. [[CrossRef](#)]
37. Singh, A.; Kett, W.C.; Severin, I.C.; Agyekum, I.; Duan, J.; Amster, I.J.; Proudfoot, A.E.I.; Coombe, D.R.; Woods, R.J. The interaction of heparin tetrasaccharides with chemokine CCL5 is modulated by sulfation pattern and pH. *J. Biol. Chem.* **2015**, *290*, 15421–15436. [[CrossRef](#)]

38. Agostino, M.; Gandhi, N.S.; Mancera, R.L. Development and application of site mapping methods for the design of glycosaminoglycans. *Glycobiology* **2014**, *24*, 840–851. [[CrossRef](#)]
39. Marcisz, M.; Huard, B.; Lipska, A.G.; Samsonov, S.A. Further analyses of APRIL/APRIL-receptor/glycosaminoglycan interactions by biochemical assays linked to computational studies. *Glycobiology* **2021**, *31*, 772–786. [[CrossRef](#)]
40. Winkler, S.; Derler, R.; Gesslbauer, B.; Krieger, E.; Kungl, A.J. Molecular dynamics simulations of the chemokine CCL2 in complex with pull down-derived heparan sulfate hexasaccharides. *Biochim. Biophys. Acta-Gen. Subj.* **2019**, *1863*, 528–533. [[CrossRef](#)]
41. Marcisz, M.; Zacharias, M.; Samsonov, S.A. Modeling protein-glycosaminoglycan complexes: Does the size matter? *J. Chem. Inf. Model.* **2021**, *61*, 4475–4485. [[CrossRef](#)] [[PubMed](#)]
42. Raghuraman, A.; Mosier, P.D.; Desai, U.R. Finding a needle in a haystack: Development of a combinatorial virtual screening approach for identifying high specificity heparin/heparan sulfate sequence(s). *J. Med. Chem.* **2006**, *49*, 3553–3562. [[CrossRef](#)] [[PubMed](#)]
43. Nivedha, A.K.; Makeneni, S.; Foley, B.L.; Tessier, M.B.; Woods, R.J. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J. Comput. Chem.* **2014**, *35*, 526–539. [[CrossRef](#)] [[PubMed](#)]
44. Guerrini, M.; Guglieri, S.; Beccati, D.; Torri, G.; Viskov, C.; Mourier, P. Conformational transitions induced in heparin octasaccharides by binding with antithrombin III. *Biochem. J.* **2006**, *399*, 191–198. [[CrossRef](#)] [[PubMed](#)]
45. Hricovíni, M.; Guerrini, M.; Bisio, A.; Torri, G.; Naggi, A.; Casu, B. Active conformations of glycosaminoglycans. NMR determination of the conformation of heparin sequences complexed with antithrombin and fibroblast growth factors in solution. *Semin. Thromb. Hemost.* **2002**, *28*, 325–334. [[CrossRef](#)]
46. Nagarajan, B.; Holmes, S.G.; Sankaranarayanan, N.V.; Desai, U.R. Molecular dynamics simulations to understand glycosaminoglycan interactions in the free- and protein-bound states. *Curr. Opin. Struct. Biol.* **2022**, *74*, 102356. [[CrossRef](#)]
47. Nagarajan, B.; Sankaranarayanan, N.V.; Desai, U.R. Perspective on computational simulations of glycosaminoglycans. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, *9*, e1388. [[CrossRef](#)]
48. Samsonov, S.A.; Pisabarro, M.T. Importance of IdoA and IdoA(2S) ring conformations in computational studies of glycosaminoglycan-protein interactions. *Carbohydr. Res.* **2013**, *381*, 133–137. [[CrossRef](#)]
49. Sankaranarayanan, N.V.; Bi, Y.; Kuberan, B.; Desai, U.R. Combinatorial virtual library screening analysis of antithrombin binding oligosaccharide motif generation by heparan sulfate 3-O-Sulfotransferase 1. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 933–941. [[CrossRef](#)]
50. Liebeschuetz, J.; Hennemann, J.; Olsson, T.; Groom, C.R. The good, the bad and the twisted: A survey of ligand geometry in protein crystal structures. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 169–183. [[CrossRef](#)]
51. Xu, D.; Esko, J.D. Demystifying heparan sulfate-protein interactions. *Annu. Rev. Biochem.* **2014**, *83*, 129–157. [[CrossRef](#)] [[PubMed](#)]
52. Whalen, D.M.; Malinauskas, T.; Gilbert, R.J.C.; Siebold, C. Structural insights into proteoglycan-shaped Hedgehog signaling. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16420–16425. [[CrossRef](#)] [[PubMed](#)]
53. Nurisso, A.; Bravo, J.; Carrupt, P.-A.; Daina, A. Molecular docking using the molecular lipophilicity potential as hydrophobic descriptor: Impact on GOLD docking performance. *J. Chem. Inf. Model.* **2012**, *52*, 1319–1327. [[CrossRef](#)]
54. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved protein-ligand docking using GOLD. *Proteins Struct. Funct. Bioinform.* **2003**, *52*, 609–623. [[CrossRef](#)]
55. Detering, C.; Varani, G. Validation of automated docking programs for docking and database screening against RNA drug targets. *J. Med. Chem.* **2004**, *47*, 4188–4201. [[CrossRef](#)] [[PubMed](#)]
56. Kwon, S.; Seok, C. CSAlign and CSAlign-Dock: Structure alignment of ligands considering full flexibility and application to protein–ligand docking. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 1–10. [[CrossRef](#)] [[PubMed](#)]
57. Greenidge, P.A.; Lewis, R.A.; Ertl, P. Boosting pose ranking performance via rescoring with MM-GBSA. *Chem. Biol. Drug Des.* **2016**, *88*, 317–328. [[CrossRef](#)]
58. Morris, G.M.; Goodsell, D.S.; Huey, R.; Olson, A.J. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293–304. [[CrossRef](#)]
59. Stefaniak, F.; Bujnicki, J.M. AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. *PLoS Comput. Biol.* **2021**, *17*, e1008309. [[CrossRef](#)]
60. Li, W.; Johnson, D.J.D.; Esmon, C.T.; Huntington, J.A. Structure of the antithrombin–thrombin–heparin ternary complex reveals the antithrombotic mechanism of heparin. *Nat. Struct. Mol. Biol.* **2004**, *11*, 857–862. [[CrossRef](#)]
61. Zhang, Q.; Cao, H.-Y.; Wei, L.; Lu, D.; Du, M.; Yuan, M.; Shi, D.; Chen, X.; Wang, P.; Chen, X.-L.; et al. Discovery of exolytic heparinases and their catalytic mechanism and potential application. *Nat. Commun.* **2021**, *12*, 1263. [[CrossRef](#)] [[PubMed](#)]
62. Holmes, S.G.; Nagarajan, B.; Desai, U.R. 3-O-Sulfation induces sequence-specific compact topologies in heparan sulfate that encode a dynamic sulfation code. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 3884–3898. [[CrossRef](#)] [[PubMed](#)]
63. Janke, J.J.; Yu, Y.; Pomin, V.H.; Zhao, J.; Wang, C.; Linhardt, R.J.; García, A.E. Characterization of heparin’s conformational ensemble by molecular dynamics simulations and nuclear magnetic resonance spectroscopy. *J. Chem. Theory Comput.* **2022**, *18*, 1894–1904. [[CrossRef](#)]
64. Cilpa, G.; Hyvönen, M.T.; Koivuniemi, A.; Riekkola, M.L. Atomistic insight into chondroitin-6-sulfate glycosaminoglycan chain through quantum mechanics calculations and molecular dynamics simulation. *J. Comput. Chem.* **2010**, *31*, 1670–1680. [[CrossRef](#)]
65. Hricovíni, M. Solution Structure of Heparin Pentasaccharide: NMR and DFT Analysis. *J. Phys. Chem. B* **2015**, *119*, 12397–12409. [[CrossRef](#)] [[PubMed](#)]

66. Hricovíni, M.; Hricovíni, M. Solution conformation of heparin tetrasaccharide. DFT analysis of structure and spin–spin coupling constants. *Molecules* **2018**, *23*, 3042. [[CrossRef](#)]
67. Pagielska, M.; Samsonov, S.A. Molecular dynamics-based comparative analysis of chondroitin and dermatan sulfates. *Biomolecules* **2023**, *13*, 247. [[CrossRef](#)]
68. Gesteira, T.F.; Pol-Fachin, L.; Coulson-Thomas, V.J.; Lima, M.A.; Verli, H.; Nader, H.B. Insights into the N-sulfation mechanism: Molecular dynamics simulations of the N-sulfotransferase domain of Ndst1 and mutants. *PLoS ONE* **2013**, *8*, e70880. [[CrossRef](#)]
69. Joseph, P.R.B.; Mosier, P.D.; Desai, U.R.; Rajarathnam, K. Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: Structural plasticity mediates differential binding interactions. *Biochem. J.* **2015**, *472*, 121–133. [[CrossRef](#)]
70. Nakamichi, Y.; Mikami, B.; Murata, K.; Hashimoto, W. Crystal structure of a bacterial unsaturated glucuronyl hydrolase with specificity for heparin. *J. Biol. Chem.* **2014**, *289*, 4787–4797. [[CrossRef](#)]
71. Zhao, B.; LiWang, P.J. Characterization of the interactions of vMIP-II, and a dimeric variant of vMIP-II, with glycosaminoglycans. *Biochemistry* **2010**, *49*, 7012–7022. [[CrossRef](#)] [[PubMed](#)]
72. Zhang, X.; Lin, L.; Huang, H.; Linhardt, R.J. Chemoenzymatic synthesis of glycosaminoglycans. *Acc. Chem. Res.* **2020**, *53*, 335–346. [[CrossRef](#)] [[PubMed](#)]
73. Zulueta, M.M.L.; Lin, S.Y.; Hu, Y.P.; Hung, S.C. Synthesis of glycosaminoglycans. In *Glycochemical Synthesis: Strategies and Applications*; Wiley Blackwell: Hoboken, NJ, USA, 2015; pp. 235–261. [[CrossRef](#)]
74. Balius, T.E.; Fischer, M.; Stein, R.M.; Adler, T.B.; Nguyen, C.N.; Cruz, A.; Gilson, M.K.; Kurtzman, T.; Shoichet, B.K. Testing inhomogeneous solvation theory in structure-based ligand discovery. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E6839–E6846. [[CrossRef](#)] [[PubMed](#)]
75. Fischer, M.; Coleman, R.G.; Fraser, J.S.; Shoichet, B.K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* **2014**, *6*, 575–583. [[CrossRef](#)]
76. Sarkar, A.; Yu, W.; Desai, U.R.; Mackerell, A.D.; Mosier, P.D. Estimating glycosaminoglycan-protein interaction affinity: Water dominates the specific antithrombin-heparin interaction. *Glycobiology* **2016**, *26*, 1041–1047. [[CrossRef](#)]
77. Liang, W.G.; Triandafillou, C.G.; Huang, T.Y.; Zulueta, M.M.L.; Banerjee, S.; Dinner, A.R.; Hung, S.C.; Tang, W.J. Structural basis for oligomerization and glycosaminoglycan binding of CCL5 and CCL3. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5000–5005. [[CrossRef](#)] [[PubMed](#)]
78. Shaw, J.P.; Johnson, Z.; Borlat, F.; Zwahlen, C.; Kungl, A.; Roulin, K.; Harrenga, A.; Wells, T.N.C.; Proudfoot, A.E.I. The X-ray structure of RANTES: Heparin-derived disaccharides allows the rational design of chemokine inhibitors. *Structure* **2004**, *12*, 2081–2093. [[CrossRef](#)]
79. Carter, W.J.; Cama, E.; Huntington, J.A. Crystal structure of thrombin bound to heparin. *J. Biol. Chem.* **2005**, *280*, 2745–2749. [[CrossRef](#)]
80. Koo, B.H.; Han, J.H.; Yeom, Y.I.I.; Kim, D.S. Thrombin-dependent MMP-2 activity is regulated by heparan sulfate. *J. Biol. Chem.* **2010**, *285*, 41270–41279. [[CrossRef](#)]
81. Wang, C.; Jin, Y.; Desai, U.R.; Yadavalli, V.K. Investigation of the heparin-thrombin interaction by dynamic force spectroscopy. *Biochim. Biophys. Acta-Gen. Subj.* **2015**, *1850*, 1099–1106. [[CrossRef](#)]
82. Cai, Z.; Yarovoi, S.V.; Zhu, Z.; Rauova, L.; Hayes, V.; Lebedeva, T.; Liu, Q.; Poncz, M.; Arepally, G.; Cines, D.B.; et al. Atomic description of the immune complex involved in heparin-induced thrombocytopenia. *Nat. Commun.* **2015**, *6*, 8277. [[CrossRef](#)] [[PubMed](#)]
83. Datta, P.; Zhang, F.; Dordick, J.S.; Linhardt, R.J. Platelet factor 4 polyanion immune complexes: Heparin induced thrombocytopenia and vaccine-induced immune thrombotic thrombocytopenia. *Thromb. J.* **2021**, *19*, 66. [[CrossRef](#)]
84. Nguyen, T.H.; Greinacher, A.; Delcea, M. Quantitative description of thermodynamic and kinetic properties of the platelet factor 4/heparin bonds. *Nanoscale* **2015**, *7*, 10130–10139. [[CrossRef](#)] [[PubMed](#)]
85. Niu, C.; Yang, Y.; Huynh, A.; Nazy, I.; Kaltashov, I.A. Platelet factor 4 interactions with short heparin oligomers: Implications for folding and assembly. *Biophys. J.* **2020**, *119*, 1371–1379. [[CrossRef](#)] [[PubMed](#)]
86. Belvedere, R.; Morretta, E.; Pessolano, E.; Novizio, N.; Tosco, A.; Porta, A.; Whiteford, J.; Perretti, M.; Filippelli, A.; Monti, M.C.; et al. Mesoglycan exerts its fibrinolytic effect through the activation of annexin A2. *J. Cell. Physiol.* **2021**, *236*, 4926–4943. [[CrossRef](#)] [[PubMed](#)]
87. Kassam, G.; Manro, A.; Braat, C.E.; Louie, P.; Fitzpatrick, S.L.; Waisman, D.M. Characterization of the heparin binding properties of annexin II tetramer. *J. Biol. Chem.* **1997**, *272*, 15093–15100. [[CrossRef](#)]
88. Shao, C.; Zhang, F.; Kemp, M.M.; Linhardt, R.J.; Waisman, D.M.; Head, J.F.; Seaton, B.A. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *J. Biol. Chem.* **2006**, *281*, 31689–31695. [[CrossRef](#)]
89. Ahmed, Y.A.; Yates, E.A.; Moss, D.J.; Loeven, M.A.; Hussain, S.A.; Hohenester, E.; Turnbull, J.E.; Powell, A.K. Panels of chemically-modified heparin polysaccharides and natural heparan sulfate saccharides both exhibit differences in binding to Slit and Robo, as well as variation between protein binding and cellular activity. *Mol. Biosyst.* **2016**, *12*, 3166–3175. [[CrossRef](#)]
90. Fukuhara, N.; Howitt, J.A.; Hussain, S.A.; Hohenester, E. Structural and functional analysis of slit and heparin binding to immunoglobulin-like domains 1 and 2 of Drosophila robo. *J. Biol. Chem.* **2008**, *283*, 16226–16234. [[CrossRef](#)]
91. Li, Z.; Moniz, H.; Wang, S.; Ramiah, A.; Zhang, F.; Moremen, K.W.; Linhardt, R.J.; Sharp, J.S. High structural resolution hydroxyl radical protein footprinting reveals an extended Robo1-heparin binding interface. *J. Biol. Chem.* **2015**, *290*, 10729–10740. [[CrossRef](#)]

92. Williams, R.V.; Huang, C.; Moremen, K.W.; Amster, I.J.; Prestegard, J.H. NMR analysis suggests the terminal domains of Robo1 remain extended but are rigidified in the presence of heparan sulfate. *Sci. Rep.* **2022**, *12*, 14769. [[CrossRef](#)] [[PubMed](#)]
93. Chang, S.C.; Mulloy, B.; Magee, A.I.; Couchman, J.R. Two distinct sites in sonic hedgehog combine for heparan sulfate interactions and cell signaling functions. *J. Biol. Chem.* **2011**, *286*, 44391–44402. [[CrossRef](#)] [[PubMed](#)]
94. Ortmann, C.; Pickhinke, U.; Exner, S.; Ohlig, S.; Lawrence, R.; Jboor, H.; Dreier, R.; Grobe, K. Correction to Sonic hedgehog processing and release are regulated by glypican heparan sulfate proteoglycans. *J. Cell Sci.* **2015**, *128*, 2374–2385. [[CrossRef](#)] [[PubMed](#)]
95. Deshauer, C.; Morgan, A.M.; Ryan, E.O.; Handel, T.M.; Prestegard, J.H.; Wang, X. Interactions of the chemokine CCL5/RANTES with medium-sized chondroitin sulfate ligands. *Structure* **2015**, *23*, 1066–1077. [[CrossRef](#)] [[PubMed](#)]
96. Dyer, D.P.; Salanga, C.L.; Volkman, B.F.; Kawamura, T.; Handel, T.M. The dependence of chemokine-glycosaminoglycan interactions on chemokine oligomerization. *Glycobiology* **2015**, *26*, 312–326. [[CrossRef](#)]
97. Ofori, F.A.; Modi, G.J.; Smith, L.M.; Cerskus, A.L.; Hirsh, J.; Blajchman, M.A. Heparan sulfate and dermatan sulfate inhibit the generation of thrombin activity in plasma by complementary pathways. *Blood* **1984**, *64*, 742–747. [[CrossRef](#)]
98. Cao, L.; Coventry, B.; Goresnik, I.; Huang, B.; Sheffler, W.; Park, J.S.; Jude, K.M.; Marković, I.; Kadam, R.U.; Verschueren, K.H.G.; et al. Design of protein-binding proteins from the target structure alone. *Nature* **2022**, *605*, 551–560. [[CrossRef](#)]
99. Huang, S.Y.; Zou, X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins Struct. Funct. Genet.* **2008**, *72*, 557–579. [[CrossRef](#)]
100. Kaufmann, K.W.; Lemmon, G.H.; Deluca, S.L.; Sheehan, J.H.; Meiler, J. Practically useful: What the Rosetta protein modeling suite can do for you. *Biochemistry* **2010**, *49*, 2987–2998. [[CrossRef](#)]
101. Sankaranarayanan, N.V.; Strelbe, T.R.; Boothello, R.S.; Sheerin, K.; Raghuraman, A.; Sallas, F.; Mosier, P.D.; Watermeyer, N.D.; Oscarson, S.; Desai, U.R. A hexasaccharide containing rare 2-O-sulfate-glucuronic acid residues selectively activates heparin cofactor II. *Angew. Chem.-Int. Ed.* **2017**, *56*, 2312–2317. [[CrossRef](#)]
102. Samsonov, S.A.; Teyra, J.; Pisabarro, M.T. Docking glycosaminoglycans to proteins: Analysis of solvent inclusion. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 477–489. [[CrossRef](#)] [[PubMed](#)]
103. Uciechowska-Kaczmarzyk, U.; Chauvot de Beauchene, I.; Samsonov, S. Docking software performance in protein-glycosaminoglycan systems. *J. Mol. Graph. Model.* **2019**, *90*, 42–50. [[CrossRef](#)] [[PubMed](#)]
104. Griffith, A.R.; Rogers, C.J.; Miller, G.M.; Abrol, R.; Hsieh-Wilson, L.C.; Goddard, W.A. Predicting glycosaminoglycan surface protein interactions and implications for studying axonal growth. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 13697–13702. [[CrossRef](#)] [[PubMed](#)]
105. Chopra, P.; Joshi, A.; Wu, J.; Lu, W.; Yadavalli, T.; Wolfert, M.A.; Shukla, D.; Zaia, J.; Boons, G.-J. The 3-O-sulfation of heparan sulfate modulates protein binding and lyase degradation. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2012935118. [[CrossRef](#)]
106. Liu, J.; Shriver, Z.; Marshall Pope, R.; Thorp, S.C.; Duncan, M.B.; Copeland, R.J.; Raska, C.S.; Yoshida, K.; Eisenberg, R.J.; Cohen, G.; et al. Characterization of a heparan sulfate octasaccharide that binds to herpes simplex virus type 1 glycoprotein D. *J. Biol. Chem.* **2002**, *277*, 33456–33467. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.